Statistica Sinica **18**(2008), 1483-1500

INFERRING INDIVIDUAL DIFFERENCES IN FMRI: FINDING BRAIN REGIONS WITH SIGNIFICANT WITHIN SUBJECT CORRELATION

Sumitra Purkayastha, Tor D. Wager and Thomas E. Nichols

Indian Statistical Institute, Columbia University and GlaxoSmithKline

Abstract: Functional magnetic resonance imaging studies answer questions about activation effects in populations of subjects. To begin with, this involves appropriate modeling of the fMRI data at the within-subject level. This is followed by extending the model to multiple subjects. There have been several attempts toward this extension, all of which have focused on inference on a single effect of interest (e.g., fMRI response for one type of working memory). However, the existing literature does not seem to say much about the relevant inferential procedures when *multiple* effects are of interest (e.g., response for four different types of working memory). In particular, the *within subject* dependence of one activation effect on another is an important issue with a multivariate repeated measures model. While most standard statistical methods regard such correlation as a nuisance, to be adjusted for and then ignored, we develop two simple and intuitive tests to make inference on the existence of such correlation. We demonstrate use of these tests by application to an fMRI study of attention switching. These tests are different not only from conventional tests for sphericity but also, more importantly, from the likelihood ratio test (LRT) of the relevant hypothesis. We also discuss what prompts us to look for tests different from the LRT.

Key words and phrases: Attention switching, functional magnetic resonance imaging, likelihood ratio test, mixed model.

1. Introduction

Functional magnetic resonance imaging studies answer questions about activation effects in populations of subjects. To begin with, this involves appropriate modeling of the fMRI data at the individual single-session level (e.g., Bullmore et al. (1996), Woolrich et al. (2001), Worsley and Friston (1995)). This is followed by extending the model to include data from multiple participants, treating participants as a random effect. There have been several attempts toward this extension by Holmes and Friston (1998), Worsley et al. (2002), Beckmann, Jenkinson and Smith (2003), Woolrich et al. (2004), and others. However, existing 'massive univariate' analysis tools such as SPM software (Friston et al. (2002), FSL (Smith et al. (2004)), or AFNI (Cox (1996)) are geared toward providing

inferences on a set of activation coefficients from a single contrast (i.e., whether voxels are 'activated'), not on dealing with covariance among multiple contrasts (i.e., whether activations among several contrasts are correlated). Some software packages provide facility for testing whether the multivariate mean of a set of activation coefficients is nonzero (e.g., using an F-test across multiple contrasts in SPM2). This is a test of the location of coefficients from several contrasts, and a significant result indicates confidence that the set of contrasts produced increases or decreases in activity on average. Here we are interested in developing methods for testing the covariance of coefficients from several contrasts. In such tests, a significant result indicates confidence that the activations across task types are correlated. In brief, we seek a test for correlations among responses evoked by different tasks, expressed over subjects. This can be regarded as a test for non-sphericity induced by correlations or dependencies among task-dependent responses at the between-subject level.

Let us illustrate this problem with a concrete example. Recently, we provided support for the idea that multiple types of attention shifting increase activity in the same voxels in parietal and posterior prefrontal cortices (Wager, Jonides, Smith and Nichols (2005a)). One natural conclusion is that the same brain regions are involved in different types of attention shifts. However, the analyses that support this conclusion ignore a very important additional source of evidence contained in the data; namely, whether individuals who show high activity in a region in one shift type also show high activity in other shift types. Testing correlations across measures of performance has been fundamental to the development of theories of intelligence and cognitive function in psychology (Dempster and Corkill (1999), Duncan et al. (1996), Kane and Engle (2003), Miyake et al. (2000), Wager, Jonides and Smith (2006) and Sylvester, Lacey, Nee, Franklin and Jonides (2005c)). The logic is that if measures are correlated, they are likely to share common underlying mental abilities. Positive correlations across a range of cognitive tasks, for example, is a primary basis for the notion of general fluid intelligence (G) (Burgess et al. (1998), Duncan, Burgess and Emslie (1995), and Duncan et al. (1996)). While measures of correlations across tasks are the 'bread and butter' of many cognitive and social scientists, they have not been widely applied to brain imaging data. Doing so would require tests of the form we propose.

In this paper, we develop a test for whether activation contrast estimates are correlated across a set of tasks. The method applies to data from N subjects on q tasks (or contrasts of interest) measured at one brain voxel. A single summary statistic provides an inferential test of the null hypothesis that activation for the q contrasts is uncorrelated. We apply the method to each voxel in the brain to generate whole-brain statistical parametric maps (SPMs) for the relatedness of

activation measures on the set of q contrasts. The statistical analysis is developed from first principles, and it avoids issues in earlier likelihood ratio tests (Neter et al. (1996)) that rely on approximations based on asymptotic assumptions. While the development of the test is motivated by application to brain imaging, it could be applied to any set of data, i.e., behavioral performance data on a set of q tasks.

The organization of the paper is as follows. In Section 2, we describe the attention shifting task design and the psychological hypotheses to be tested. In Section 3, we describe our model, the data characteristics, and the statistical formulation of the problem. In Section 4, we describe the inferential procedures for the tests. In Section 5, we present results from the attention-switching data set. In Section 6, we discuss the results and the tests, and highlight issues for future work.

2. Task Design and Psychological Hypotheses

While the main purpose of this paper is to explicate a new statistical method for analyzing covariance structures in brain imaging data, we illustrate the usefulness of the method by applying it to an empirical dataset to address questions of theoretical interest in the domain of cognitive control. Previous work on task switching and other executive functions has examined correlations in performance across different types of executive tasks (Miyake et al. (2000), Salthouse et al. (1998), Wager, Jonides and Smith (2006) and Ward, Roberts and Phillips (2001)), and found that response latencies to switch attention from one object or object feature to another are among the more reliable of the available measures of cognitive control.

In a typical switching paradigm, an experimental participant must perform a series of speeded judgments on some attribute of an object (e.g., the color, shape, size, orientation, numerosity, etc.). Cues before or during the presentation of each stimulus in the series instruct the participant which attribute to judge. On 'no-switch ' trials, the attribute judged (e.g., shape) is the same as on the previous trial (e.g., shape), whereas on 'switch' trials, the attribute judged (e.g., shape) is different that the one judged on the previous trial (e.g., orientation). A large number of studies have documented reliable increases in reaction time and switch costs for 'switch ' trials compared with 'no-switch' trials. Recently, we reported modest but quite reliable correlations among four types of switching tasks (Wager, Jonides and Smith (2006)). Participants either switched which of two attributes was judged (shape or orientation of a composite stimulus) or which of two objects the judgment was applied to. We refer to these as 'attribute' and 'object' switching, respectively. Furthermore, in some blocks of trials, objects were visible on-screen throughout the trials, whereas in others, objects disappeared after an initial learning period (usually 1-2 s) and all judgments

were made from memory. We refer to these as 'external' switches and 'internal' switches of attention, respectively. Thus, four types of attention switching were crossed in a 2×2 factorial design, with switching type (Attribute or Object) crossed with locus of representation (External or Internal).

We used this task in a recent fMRI study using this paradigm (Wager et al. (2005a)). Participants ($N = 39$) performed switch and non-switch trials of each type, which were pseudo-randomly intermixed in a rapid event-related design. The brain measure of interest for each switch type was fMRI blood-oxygen level dependent (BOLD) activation to switch trials in each type, contrasted with matching non-switch trials in each type. More details about the acquisition and analysis methods can be found in Wager et al. (2005a). Here, we focus on analysis of covariances among activation coefficients for the four switch types.

The procedures we develop test the omnibus null hypothesis of no correlations among tasks. They may reveal brain regions associated with common mechanisms for different types of task switch in a way that simply co-localizing activations in different switch types cannot. The logic is that if a brain area encodes a mechanism common across tasks, and there are reliable individual differences in how strongly this mechanism is engaged, then a participant who activates more on one task should also activate more on the others. This information can only be obtained by analyzing covariances among task activation coefficients.

Importantly, analyzing covariances in brain activity among tasks may provide new information that cannot be obtained by analyzing covariances in task performance alone. Correlations among performance across cognitive control tasks, while often reliable, have been notoriously low (Rabbitt and Lowe (2000) and Robbins et al. (1998)). This is undoubtedly so partly because performance measures on switching and other control tasks reflect multiple underlying processes, each presumably implemented in different brain mechanisms, as illustrated conceptually in Figure 1 for task switching. Some underlying processes might be common to a set of tasks (top left) and will influence brain activation across tasks in the same way. In switching, for example, good performance (low switch costs) may be obtained by efficiently ignoring irrelevant perceptual information, efficiently applying stimulus-response mapping rules, or maintaining a strong task set, among other factors (Mayr and Kliegl (2000), and Rubinstein, Meyer and Evans (2001)). If individuals are consistently good or poor at these processes, any of them might constitute a common factor. Other mental processes might be task-specific (bottom left) and either unique to one task type (as shown) or common to a subset of tasks. For example, processes related to mental imagery might be shared only by internal switching (when imagery is likely to be used to guide choices).

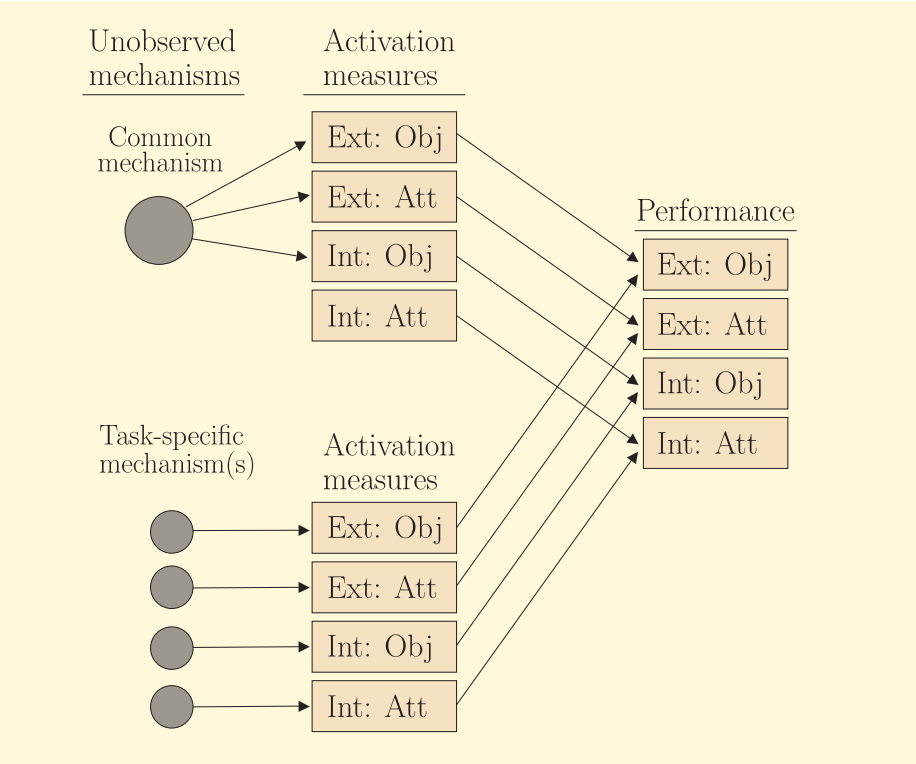

Figure 1. Schematic diagram illustrating how multiple mental processes (circles) may influence performance. The diagram illustrates common factors (top) that influence measured brain activity in different tasks (center boxes) in the same way, and task-specific factors (bottom) that influence brain activity only for one task or a subset of tasks. Behavioral performance (right) is a composite measure reflecting the combined influences of both common and task-specific factors. For this reason, examining correlations in brain activity across tasks may yield more interpretable results closer to the underlying mental processes than would examining correlations in behavioral performance.

Overall behavioral performance (Figure 1, right) thus reflects multiple, potentially independent factors. To the degree that different switching tasks differentially load on these factors, correlations in performance measures will be low. However, activation in a particular region of the brain is more likely to reflect engagement of one particular cognitive process. Thus, testing correlations across tasks in the brain may reveal the loci of common processes before they are mixed with variability due to other factors.

3. Model, Data and Statistical Formulation of the Problem

We work with a special case of the two-level linear mixed-effects model (e.g., Beckmann, Jenkinson and Smith (2003), Woolrich et al. (2004)). For a single voxel and for the kth ($k = 1, \ldots, N$) individual: (a) the $n \times 1$ vector of fMRI observations is denoted by Y_k, (b) the corresponding vector β_k of activation coefficients is of size $p \times 1$, (c) the corresponding design matrix, denoted X, is of size $n \times p$. For our particular data-set, $n = 1,440, N = 39$, and $p = 48$. We adopt the random-effects model (cf., Laird and Ware (1982))

$$\begin{aligned} &Y_k = X\beta_k + \epsilon_k,\ \epsilon_k \sim \mathrm{N}_n(0, \sigma^2 I_n),\ \ k = 1, \ldots, N, \\ &\beta_k = \beta + \delta_k,\ \ \delta_k \sim \mathrm{N}_p(0, \Sigma), \\ &\text{the errors } \epsilon_1, \ldots, \epsilon_N, \delta_1, \ldots, \delta_N \text{ are independent.} \end{aligned} \tag{3.1}$$

The first equation assumes that the data have been pre-whitened; the second equation expresses variation of the activation coefficients as a function of k. Notice that the design matrix X does not depend on k and that the variance σ^2 also does not depend on k. However, σ and Σ **do** depend on the voxel under consideration.

Our goal is to be able to test whether the dispersion matrix of a set of contrasts of the vector β_k of activation parameters is diagonal. Thus, if C is a given $q \times p$ matrix ($q = 4$, in our data-set) whose rows give the contrasts under consideration, we want to test if the dispersion matrix $C\Sigma C^{\mathrm{T}}$ of $C\beta_k$ is diagonal.

Notice from the first equation of (3.1), that the ordinary least squares estimate of β_k and the corresponding dispersion matrix are $\hat{\beta}_k = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y_k$ and $\mathrm{Var}(\hat{\beta}_k) = \sigma^2(X^{\mathrm{T}}X)^{-1}$, respectively. Consequently, an estimate of $C\beta_k$ is given by $Z_k \stackrel{\mathrm{def}}{=} C\hat{\beta}_k = DY_k$, where $D \stackrel{\mathrm{def}}{=} C(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$. Write $\mathrm{Var}(Z_k) = \sigma^2 U$, with $U = C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}$, and let $\gamma_k = C\beta_k, \gamma = C\beta$. We pre-multiply both sides of the *first equation* of (3.1) by D, and both sides of the *second equation* of (3.1) by C, to obtain

$$\begin{aligned} &Z_k = \gamma_k + \eta_k,\ \ \eta_k \stackrel{\mathrm{def}}{=} D\epsilon_k \sim \mathrm{N}_q(0, \sigma^2 U),\ k = 1, \ldots, N, \\ &\gamma_k = \gamma + \zeta_k,\ \ \zeta_k \stackrel{\mathrm{def}}{=} C\delta_k \sim \mathrm{N}_q(0, \Gamma),\ \Gamma \stackrel{\mathrm{def}}{=} C\Sigma C^{\mathrm{T}}, \\ &\eta_1, \ldots, \eta_N, \zeta_1, \ldots, \zeta_N \text{ are independent.} \end{aligned} \tag{3.2}$$

Notice that the model specified by (3.2) is not identifiable, as the distribution of $(Z_1, \ldots, Z_N)$ under $(\gamma, \sigma_1,\ \sigma_2^2 U)$ is same as that of $(Z_1, \ldots, Z_N)$ under $(\gamma, \sigma_2, \sigma_1^2 U)$. So, we assume in (3.2) that σ *is known and equals* σ_0. In other words, $Z_k \overset{i.i.d.}{\sim} \mathrm{N}_q(\gamma, \sigma_0^2 U + \Gamma),\ k = 1, \ldots, N$. The assumption that σ is known is

reasonable, as an accurate estimate of σ can be obtained. We discuss in Section 4.3 how we choose and fix σ_0.

We formulate now our question in terms of Γ as

$$\mathrm{H}_0 : \Gamma \text{ is diagonal} \quad \text{against} \quad \mathrm{H}_1 : \Gamma \text{ is non-diagonal.} \tag{3.3}$$

4. Statistical Analysis: Two Simple and Intuitive Tests for (3.3)

We suggest in the appendix that finding even a tractable expression for the LRT for (3.3) is quite formidable, not to speak of the difficulty of finding its asymptotic distribution. Observe from $Z_k \overset{i.i.d.}{\sim} \mathrm{N}_q(\gamma, \sigma_0^2 U + \Gamma)$, $k = 1, \ldots, N$, that conventional tests for non-sphericity (see, e.g., Section 10.7 of Anderson (2003)) are not applicable for testing (3.3). We need to get rid of U. In view of these facts, in what follows we develop two simple and intuitively plausible tests for the problem at (3.3).

Write $U = ((u_{ij}))$, $\Gamma = ((\gamma_{ij}))$, and take $\Lambda = \sigma_0^2 U + \Gamma = ((\lambda_{ij}))$, say, where $\lambda_{ij} = \sigma_0^2 u_{ij} + \gamma_{ij}$. We can write (3.3) as

$$\mathrm{H}_0 : \lambda_{ij} = \sigma_0^2 u_{ij} \text{ for all } i < j \qquad \text{against} \qquad \mathrm{H}_1 : \lambda_{ij} \neq \sigma_0^2 u_{ij} \text{ for some } i < j. \tag{4.1}$$

It should be noted that here $\#\{(i,j) : 1 \leq i < j \leq q\} = q(q-1)/2 = 6$. In other words, H_0 is associated with 6 hypotheses, corresponding to the elements in the upper (or lower) triangle of Γ.

Let us recall the following result about asymptotic distribution of the entries of a sample covariance matrix, when samples are being drawn from a multivariate normal distribution.

Theorem.(Anderson (2003, p.87)) *Let* $A_N \overset{\text{def}}{=} \sum_{\alpha=1}^{N} (X_\alpha - \bar{X}_N)(X_\alpha - \bar{X}_N)^\tau$, *where* $X_1, X_2, \ldots$ *are i.i.d.* $\mathrm{N}_p(\mu, \Psi)$ *and* $N\,\bar{X}_N \overset{\text{def}}{=} \sum_{\alpha=1}^{N} X_\alpha$. *Then the limiting distribution of* $B_N \overset{\text{def}}{=} (\sqrt{N-1})^{-1}[A_N - (N-1)\Psi] = ((b_{ij,N}))$ *is normal with mean* 0 *and covariances* $\mathrm{E}(b_{ij,N}\, b_{kl,N}) = \psi_{ik}\psi_{jl} + \psi_{il}\psi_{jk}$.

Corollary. *By taking* $i = k$, $j = l$, *in the theorem above, it follows that the asymptotic distribution of* $b_{ij,N}$ *is univariate normal with mean* 0 *and variance* $\psi_{ii}\psi_{jj} + \psi_{ij}^2$.

We make use of the theorem and the corollary to propose two simple and intuitively plausible tests for (3.3). Let us define the $q \times q$ matrices $S_N = ((s_{ij,N}))$ and $T_N = ((t_{ij,N}))$ by $(N-1)\, S_N \overset{\text{def}}{=} \sum_{\alpha=1}^{N} (Z_\alpha - \bar{Z}_N)(Z_\alpha - \bar{Z}_N)^{\mathrm{T}}$, where $N\,\bar{Z}_N \overset{\text{def}}{=}$

$\sum_{\alpha=1}^{N} Z_\alpha$, and $T_N \stackrel{\text{def}}{=} \sqrt{N-1}\,[S_N - \Lambda]$. It should be emphasized that these matrices refer to one voxel, and calculations are repeated for each one.

4.1. The first test for (3.3)

The asymptotic distribution of $t_{ij,N} = \sqrt{N-1}\,(s_{ij,N} - \lambda_{ij}) = \sqrt{N-1}\,(s_{ij,N} - \sigma_0^2 u_{ij} - \gamma_{ij})$ is univariate normal with mean 0 and variance $\lambda_{ii}\lambda_{jj} + \lambda_{ij}^2 = (\sigma_0^2 u_{ii} + \gamma_{ii})(\sigma_0^2 u_{jj} + \gamma_{jj}) + (\sigma_0^2 u_{ij} + \gamma_{ij})^2$, and a consistent estimate of $\lambda_{ii}\lambda_{jj} + \lambda_{ij}^2$ is $s_{ii,N}\, s_{jj,N} + s_{ij,N}^2$. Also, under H_0, $\lambda_{ij} = \sigma_0^2 u_{ij}$. Therefore, under H_0,

$$v_{ij,N} \stackrel{\text{def}}{=} \frac{\sqrt{N-1}\,(s_{ij,N} - \sigma_0^2 u_{ij})}{\sqrt{s_{ii,N}\, s_{jj,N} + s_{ij,N}^2}} \xrightarrow{d} \mathrm{N}(0,1) \text{ as } N \to \infty. \tag{4.2}$$

This enables us to construct a test for

$$\mathrm{H}_0^{i,j} : \lambda_{ij} = \sigma_0^2 u_{ij} \quad \text{against} \quad \mathrm{H}_1^{i,j} : \lambda_{ij} \neq \sigma_0^2 u_{ij}, \tag{4.3}$$

for every fixed $i < j$. We implement the test for every pair (i, j) with $i < j$ and employ Bonferroni's inequality to test (3.3). The number of hypotheses of the type (4.3) associated with (3.3) is 6 and so application of Bonferroni's inequality is not too conservative in this context. The details are given below.

Fix α. We reject H_0 if

$$T_{N,1} \stackrel{\text{def}}{=} \max_{i<j} |v_{ij,N}| = \max_{i<j} \left| \frac{\sqrt{N-1}\,(s_{ij,N} - \sigma_0^2 u_{ij})}{\sqrt{s_{ii,N}\, s_{jj,N} + s_{ij,N}^2}} \right| > z_{1-\frac{\alpha}{12}}, \tag{4.4}$$

where z_p is the pth quantile of standard normal distribution. For example, $z_{1-\alpha/12} = 2.6383$ (3.1440), when $\alpha = 0.05\,(0.01)$.

4.2. The second test for (3.3)

The previous test uses the maximum of the off-diagonal elements of S_N, suitably normalized, while the next test combines these elements to provide a test for (3.3). Let

$$W_N = \sqrt{N-1} \begin{pmatrix} s_{12,N} - \sigma_0^2 u_{12} \\ s_{13,N} - \sigma_0^2 u_{13} \\ s_{14,N} - \sigma_0^2 u_{14} \\ s_{23,N} - \sigma_0^2 u_{23} \\ s_{24,N} - \sigma_0^2 u_{24} \\ s_{34,N} - \sigma_0^2 u_{34} \end{pmatrix}, \tag{4.5.1}$$

and, by using the simplified notation s_{ij} in place of $s_{ij,N}$,

$$\hat{\Delta}_N = \begin{pmatrix} s_{11}s_{22} & s_{11}s_{23} & s_{11}s_{24} & s_{12}s_{23} & s_{12}s_{24} & s_{13}s_{24} \\ & s_{11}s_{33} & s_{11}s_{34} & s_{12}s_{33} & s_{12}s_{34} & s_{13}s_{34} \\ & & s_{11}s_{44} & s_{12}s_{34} & s_{12}s_{44} & s_{13}s_{44} \\ & & & s_{22}s_{33} & s_{22}s_{34} & s_{23}s_{34} \\ & & & & s_{22}s_{44} & s_{23}s_{44} \\ & & & & & s_{33}s_{44} \end{pmatrix} + \begin{pmatrix} s_{12}^2 & s_{13}s_{12} & s_{14}s_{12} & s_{13}s_{22} & s_{14}s_{22} & s_{14}s_{23} \\ & s_{13}^2 & s_{14}s_{13} & s_{13}s_{23} & s_{14}s_{23} & s_{14}s_{33} \\ & & s_{14}^2 & s_{13}s_{24} & s_{14}s_{24} & s_{14}s_{34} \\ & & & s_{23}^2 & s_{23}s_{24} & s_{24}s_{33} \\ & & & & s_{24}^2 & s_{24}s_{34} \\ & & & & & s_{34}^2 \end{pmatrix}. \tag{4.5.2}$$

Proceeding essentially along the same lines as in Section 4.1, it can be seen that, under H_0, the asymptotic distribution of (as $N \to \infty$)

$$T_{N,2} \stackrel{\text{def}}{=} W_N^{\mathrm{T}} \hat{\Delta}_N^{-1} W_N \tag{4.5.3}$$

is chi-square with $d = 6$ degrees of freedom. This fact enables us to construct a test for (3.3). The details are given below.

Fix α. We reject H_0, if

$$T_{N,2} > \chi_{1-\alpha,6}, \tag{4.5.6}$$

where $\chi_{p,\nu}$ is the pth quantile of χ^2 distribution with ν degrees of freedom. For example, $\chi_{1-\alpha/6} = 12.5916\,(16.8119)$ when $\alpha = 0.05\,(0.01)$.

4.3. Choice of σ_0^2

We have assumed earlier (cf., last but one paragraph of Section 3) that σ is known to be σ_0. We describe below our choice of σ_0.

Consider the maximum likelihood estimate of σ^2 from each of the models $Y_k = X\beta_k + \epsilon_k,\ k = 1, \ldots, N$, where $\epsilon_1, \ldots, \epsilon_N$ are i.i.d. $N_n(0, \sigma^2 I_n)$ errors. This estimate is given by (cf. (3.1))

$$\sigma_{k,0}^2 = \frac{Y_k^{\mathrm{T}}(I_n - H)Y_k}{n},\ \ H \stackrel{\text{def}}{=} X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}},\ \ k = 1, \ldots, N. \tag{4.7.1}$$

We choose

$$\sigma_0^2 = \frac{\sum_{k=1}^N \sigma_{k,0}^2}{N} = \frac{\sum_{k=1}^N Y_k^{\mathrm{T}}(I_n - H)Y_k}{Nn} \tag{4.7.2}$$

Thus, the separate estimates of σ^2 are averaged to get σ_0^2. This is expected to be a fairly accurate estimate of σ^2 as n is large. In passing, we note that this

choice of σ^2 makes use of the Y_k's and not the Z_k's. We also note the possible dependence of σ_0 on the voxels.

4.4. Correction for multiple comparisons

The test statistics described above correct for multiple tests on covariances within a single voxel. For neuroscientific inference in cases where the technique is applied over many voxels, additional spatial correction for multiple comparisons is desirable. Such a correction using False Discovery Rate (FDR; Genovese, Lazar and Nichols (2002)) control may be easily applied to p-values generated from our test statistics. However, statistical power decreases dramatically both with the number of tasks and the number of voxels tested. Alternatively, if anatomically specific hypotheses are generated a priori (e.g., based on meta-analysis of previous work; see Wager, Jonides and Reading (2004)), less stringent correction may be imposed.

Here, we were primarily interested in demonstrating the technique, and we thus used a somewhat lenient threshold that would provide high-enough sensitivity to compare results with those expected from previous literature, restricting our interpretations to regions about which a priori anatomical hypotheses were available. Thus, our results are thresholded at $p < 0.01$ (corrected for multiple comparisons across tasks, but not across space) with an extent threshold of 20 contiguous voxels. However, we note that for the first test statistic, FDR correction across space at $q < 0.05$ yielded a threshold comparable to that employed here ($p < 0.012$), but there were no significant regions correcting across both space and pairs of tasks. For the second test, $q < 0.05$ FDR control yielded a threshold of $p < 0.0001$, exceeded by 63 voxels.

We note, in passing, that the FDR correction procedure takes into account correlations among parameter estimates across space, and results exist that provide valid FDR control with increased sensitivity when positive dependence is assumed (as is typical of fMRI data).

We conclude Section 4 by noting the following.

(a) The stability of covariance estimates relies both on the number of observations per participant ($n = 1,440$ in our experiment) and the number of participants ($N = 39$ in our study). Because the degrees of freedom at the second level is typically much smaller than at the first level, N will typically be the limiting condition for stability of contrast estimates. Whereas the test is valid for smaller samples (e.g., $N = 10-20$, currently typical of neuroimaging studies), power is expected to increase greatly as N increases. We recommend $N \approx 40$ as a minimum sample size for tests on second-level covariance.

(b) If q is moderately large, then use of Bonferroni's inequality in the first test may be conservative. Also, in the second test, we are effectively testing for the mean of a $q(q-1)/2 = 6$-dimensional approximately normal variable, with sample size $N = 39$. This may suffer from the curse of dimensionality if q is large; thus, while the test remains valid for $q > 4$, power is expected to decrease with increasing q, and it may be advisable to test lower q, particularly if only small sample sizes (e.g., $N < 100$ with $n > 500$) are available.

5. Results and Interpretation

As shown in Figure 2, we observed multiple regions in which the null hypothesis of no covariance across switch types was rejected. As the main focus of this paper is on the development of the statistical tests, we only briefly interpret these results and note how they may be of potential use in studying task switching and other psychological phenomena. Figure 2A shows maps of significant regions for test statistic $T_{N,1}$ (see (4.4)), and Figure 2B shows maps of significant regions for test statistic $T_{N,2}$ (see (4.5.3)). We note that the tests have different sensitivity to different covariance patterns, but they agree broadly in many regions of the brain. Four such regions are shown here; reading from left to right in Figure 2, they are right anterior prefrontal cortex (aPFC), left lateral orbitofrontal cortex/inferior convexity (lOFC), cuneus, and posterior cingulate cortex (PCC). The two frontal regions play roles in memory retrieval and selection of information in memory (Badre et al. (2005), Nee, Wager and Jonides (2007) and Schacter et al. (1997)), both of which are thought to be required for successful task switches (Mayr and Kliegl (2000)). The cuneus and precuneus are consistently activated in task switching (Wager, Jonides and Reading (2004)) and other executive functions (Wager and Smith (2003)). The posterior cingulate is important for aspects of memory retrieval and responds to complex visual cues (Mesulam et al. (2001) and Pandya, Van Hoesen and Mesulam (1981)).

We examined the patterns of correlations among switch types in these regions to interpret the results. Rejection of the null hypothesis could result from general "positive manifold", or positive correlations among all task types, suggesting that the region might be activated by a mechanism common to all switch types. Alternatively, the region could be activated by a mechanism shared by only some switch types, indicated by positive correlations among a subset of switch types. Finally, negative correlations could suggest strategic tradeoffs; participants might be biased toward using a strategy such as rehearsal of specific response assignments in one switch type (e.g., attribute switches), which may help performance on that switch type but hurt performance on others.

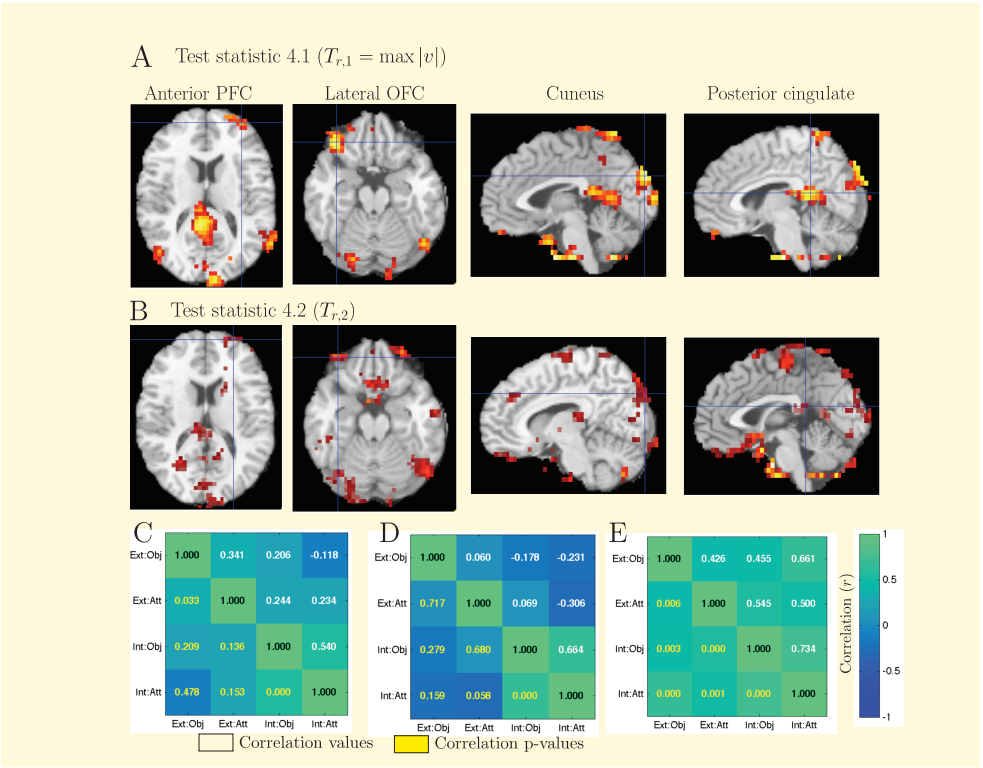

Figure 2. (A) Four regions showing significant results ($p < 0.01$) corrected across tasks and 20 contiguous voxels) in the first test (Section 4.1). The regions are, from left to right, anterior prefrontal cortex (aPFC), lateral orbitofrontal cortex (lOFC), cuneus, and posterior cingulate (PCC). (B) The same four regions also showed significant results in the second test (Section 4.2), though we note that the tests are somewhat differentially sensitive to different covariance patterns. (C) Correlations in aPFC. Correlation values are given in the upper triangle, and corresponding p-values in the lower triangle, in yellow type. The color map indexes correlation values ranging from -1 (dark blue) to 1 (bright green). Ext: External switches, stimuli visible. Int: internal switches, stimuli held in memory. Obj: Switches among objects. Att: Switches among attributes (i.e., features) of objects (e.g., shape vs. orientation).

Figure 2C shows correlations in the aPFC. Correlation values for the four switch types are in the upper triangle of the matrix, and corresponding p-values are in the lower triangle. Correlations are strongest between the two external switch types ($r = 0.34$) and the two internal switch types ($r = 0.54$), suggesting that this region might reflect processes that are engaged in both external and internal switches, but occur at a later stage of processing than other ex-

ternal or internal-specific factors. This pattern most closely mirrors the correlations in performance overall, which are strongest within external/internal and object/attribute types, and weakest across types (Wager, Jonides and Smith (2006)). This is consistent with a late-stage monitoring mechanism that tracks performance and makes adjustments, though there are other possible interpretations.

Figure 2D shows correlations in lOFC. A moderately strong correlation ($r = 0.664$) is observed in lOFC between the two internal switch types, but not for the others, suggesting that this region is recruited by a mechanism relatively unique to switching attention in working memory. Figure 2E shows correlations in PCC, though the cuneus showed the same pattern. In this region we observed general positive manifold, suggesting that all task types are intercorrelated in this region, and it may thus contain a mechanism, such as orienting to the visual stimulus, that is common across switch types and shows consistent individual differences.

Interestingly, these regions overlap with those thought to be involved in cognitive components of task switching, but notably absent are the canonical lateral prefrontal and parietal regions consistently activated in task switches (Wager, Jonides and Reading (2004)) and sometimes more generally referred to as the "attention network". This may be because the engagement of those regions is influenced by multiple lower-level factors that are differentially involved in different switch types. Each independent component process adds noise, so that brain networks that respond to the task are not necessarily the ones that show correlated activation across tasks. This argument highlights the conceptual difference between measuring overlap in activation in a set of contrasts and testing covariance among those contrasts, and suggests that tests on covariance may be useful in identifying component processes. The tests developed here represent one step in beginning to develop models of component processes, though there is much more work to be done to develop this idea. Here, we use these results merely to demonstrate the different kinds of patterns that may be identified, and we leave the attempt to provide an integrated account of processes underlying task switching for future work.

6. Conclusions and Future Work

Future work may build on the present methods to provide more tools for interpreting correlation matrices within voxels in which a significant omnibus result is obtained (i.e., using the methods developed here). Pattern hypothesis tests are a particularly attractive alternative (Steiger (1980) and Steiger (2005)). Factor-analysis methods may also be used, but the pattern hypothesis test provides a specific mechanism for testing particular patterns of intercorrelations among

tasks specified a priori, and thus offers a natural extension of the omnibus test developed here.

Second, the tests we develop may aid in identifying regions for subsequent confirmatory structural equation modeling (SEM) that could model multiple mental processes implemented in different brain networks (Mechelli et al. (2002) and Penny et al. (2004)). Conventionally, SEM has been applied to the covariance matrix encoding correlations among brain regions. However, it could also be applied to model inter-task correlations within a region of interest such as those identified in the present work.

Third, a limitation of the current work is that correlations are highly susceptible to outliers. Relationships between continuous variables can be influenced by even a few points that violate the i.i.d. assumption, particularly when they fall at the extremes of the distributions. Violations of the i.i.d. assumption can thus create both false positives and decrease sensitivity, and such considerations have spurred the development of robust estimators of covariance (e.g., Rousseeuw (1984)). Also, misregistration and/or variability in normalization to a group brain template is likely to have a significant impact on the estimation of second order effects such as correlations. Therefore, it would be worthwhile to investigate how robust the analysis is to misregistration. The use of robust regression techniques is particularly promising with neuroimaging data because of the difficulty of checking assumptions and dealing with outliers in the "massive univariate" modeling framework (Wager, Keller, Lacey and Jonides (2005b)). Future work may provide robust and nonparametric versions of the test statistics here.

Finally, there are several additional developments of the statistical model that could be pursued. First, the tests could be generalized to situations allowing for non-orthogonal designs, accounting for correlation induced by the first level design matrix. Second, Bayesian methods might be developed for the problem studied in this paper.

Acknowledgement

We would like to thank Professor K.J. Friston and Professor J. K. Ghosh for helpful comments and discussion, and two anonymous reviewers for their critical feedback. We also thank the Department of Biostatistics, University of Michigan, Ann Arbor, U.S.A., for providing facilities and resources for carrying out part of this work.

Appendix

None of the tests developed in Section 4 is the likelihood ratio test (LRT). We indicate below why finding even a tractable expression for the LRT seems formidable, not to speak of the difficulty of finding its asymptotic distribution.

Denote the likelihood function by $L(\gamma,\Gamma) \equiv L(\gamma,\Gamma|Z_1,\ldots,Z_N)$. It is given by (cf. the paragraph immediately preceding the beginning of Section 4.1)

$$L(\gamma,\Gamma) = \frac{1}{(2\pi)^{\frac{Nq}{2}}} \frac{1}{\mid \sigma_0^2 U + \Gamma \mid^{\frac{N}{2}}} \times \exp\left(-\frac{1}{2} \sum_{k=1}^{N} (Z_k - \gamma)^{\mathrm{T}} (\sigma_0^2 U + \Gamma)^{-1} (Z_k - \gamma) \right).$$

For obtaining the LRT, we have to maximize $L(\gamma,\Gamma|Z_1,\ldots,Z_N)$ both when Γ is arbitrary and when it is diagonal. For obtaining the denominator of the LRT, we have to consider set of all non-negative definite (n.n.d.) Γ. Notice that the set of matrices of the form $\sigma_0^2 U + \Gamma, \Gamma$ n.n.d is same as the set of p.d. matrices which are "bounded below" by $\sigma_0^2 U$.

It can be seen that to obtain a tractable expression of the LRT, one has to proceed along the following two steps (in principle):

(a) minimize $\log \mid \det \Sigma \mid + \mathrm{trace}(\Sigma^{-1} S)$ subject to $\Sigma \geq \Sigma_0$, i.e., $\Sigma - \Sigma_0$ is n.n.d., where S is an observed covariance matrix;

(b) minimize $\log \mid \det \Sigma \mid + \mathrm{trace}(\Sigma^{-1} S)$ subject to $\Sigma \geq \Sigma_0$ and $\Sigma - \Sigma_0$ is diagonal.

The problem in (a) has been addressed by several authors: by Michael Perlman in particular, in his Ph.D. thesis (1967) (Perlman (2006)) and by De Leeuw (2006). The solution is described below.

(1) The problem in (a), with $\Theta = \Sigma_0^{-1/2} \Sigma \Sigma_0^{-1/2}$ and $T = \Sigma_0^{-1/2} S \Sigma_0^{-1/2}$ can be seen to be equivalent to the following: minimize $\log \mid \det \Theta \mid + \mathrm{trace}(\Sigma^{-1} T)$ subject to $\Theta - I$ is n.n.d..

(2) Suppose that $T = K \Lambda K^{\mathrm{T}}$ is any spectral decomposition of T. Define the new variables $\Xi = K \Theta K^{\mathrm{T}}$. Let $\Xi = L \Omega L^{\mathrm{T}}$ be any spectral decomposition of Ξ.

(3) After some algebra, it is possible to show that the solution to the problem in (a) is given by $\hat{\Sigma} = \Sigma_0^{1/2} K^{\mathrm{T}} \hat{\Omega} K \Sigma_0^{1/2}$, where $\hat{\Omega}$ is the diagonal matrix with entries $\max(\lambda_i, 1)$.

The above steps take care of the denominator of the LRT. It is clear that finding an expression for the numerator is even more difficult, if not impossible.

To get a feel for why finding the asymptotic distribution of the LRT, whether or not we succeed in obtaining a neat expression for it, is difficult, let us consider the following univariate scenario. Suppose $X_1,\ldots,X_N$ are i.i.d. $\mathrm{N}(\mu_0,\sigma^2)$. We wish to test the hypothesis $\mathrm{H}_0 : \sigma \geq \sigma_0$ against $\mathrm{H}_1 : \sigma < \sigma_0$. It can be seen easily that the likelihood (appearing in the numerator of the LRT) is maximized at σ_0 with positive probability. This means the LRT will have a point mass. This also "indicates" that the asymptotic distribution of the LRT "may" not be χ^2.

The arguments above indicate that in our case we should not expect to get a chi-square distribution as the asymptotic distribution of the LRT. One source of the problem is that the parameter spaces, corresponding to both H_0 and $H_0 \cup H_1$, have well-defined "boundaries" and so the maximum of the likelihood will fall on these boundaries with positive probability. The asymptotics is, therefore, going to be non-standard. There is a huge literature on such non-standard asymptotics of LRT. Two important references are Chernoff (1954) and Self and Liang (1987).

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Third edition. Wiley, New Jersey.

Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z. and Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* **47**, 907-918.

Beckmann, C. F., Jenkinson, M. and Smith, S. M. (2003). General multilevel linear modeling for group analysis in fMRI. *NeuroImage* **20**, 1052-1063.

Bullmore, E., Brammer, M., Williams, S.C.R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R. and Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine* **35**, 261-277.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H. and Wilson, B. A. (1998). The ecological validity of tests of executive function. *J. Internat. Neuropsychological Soc.* **4**, 547-558.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573-578.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162-173.

Dempster, F. N. and Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica* **101**, 395-416.

Duncan, J., Burgess, P. and Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia* **33**, 261-268.

Duncan, J., Emslie, H., Williams, P., Johnson, R. and Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology* **30**, 257-303.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G. and Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: theory. *NeuroImage* **16**, 465-483.

Genovese, C. R., Lazar, N. A. and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15**, 870-878.

Holmes, A. and Friston, K. (1998). Generalisability, random effects and population inference. Fourth International Conference on Functional Mapping of the Human Brain, NeuroImage, Vol. 7, p. S754.

De Leeuw, J. (2006). Personal communication.

Kane, M. J. and Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *J. Experimental Psychology: General* **132**, 47-70.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

Mayr, U. and Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *J. Experimental Psychology-Learning Memory and Cognition* **26**, 1124-1140.

Mechelli, A., Penny, W. D., Price, C. J., Gitelman, D. R. and Friston, K. J. (2002). Effective connectivity and intersubject variability: using a multisubject network to test differences and commonalities. *NeuroImage* **17**, 1459-1469.

Mesulam, M. M., Nobre, A. C., Kim, Y. H., Parrish, T. B. and Gitelman, D. R. (2001). Heterogeneity of cingulate contributions to spatial attention. *NeuroImage* **13**, 1065-1072.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognit Psychol* **41**, 49-100.

Nee, D. E., Wager, T. D. and Jonides, J. (2007). A meta-analysis of neuroimaging studies of interference resolution. *Cogn. Affect. Behav. Neurosci.* in press.

Neter, J., Kutner, M. H., Wasserman, W. and Nachtsheim, C. J. (1996). *Applied Linear Statistical Models*. 4th edition. McGraw-Hill/Irwin.

Pandya, D. N., Van Hoesen, G. W. and Mesulam, M. M. (1981). Efferent connections of the cingulate gyrus in the rhesus monkey. *Exp. Brain Res.* **42**, 319-330.

Penny, W. D., Stephan, K. E., Mechelli, A. and Friston, K. J. (2004). Modelling functional integration: a comparison of structural equation and dynamic causal models. *NeuroImage*, **23** Suppl 1, S264-274.

Perlman, M. (2006). Personal communication.

Rabbitt, P. and Lowe, C. (2000). Patterns of cognitive ageing. *Psychol. Res.* **63**, 308-316.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., et al. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. *J. Int. Neuropsychol. Soc.* **4**, 474-490.

Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79**, 871-880.

Rubinstein, J. S., Meyer, D. E. and Evans, J. E. (2001). Executive control of cognitive processes in task switching. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 763-797.

Salthouse, T. A., Fristoe, N., McGuthry, K. E. and Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychol. Aging* **13**, 445-461.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M. and Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. *NeuroImage* **6**, 259-269.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605-610.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. and Johansen-Berg, H. et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23** Suppl 1, S208-219.

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research* **15**(3), 335-352.

Steiger, J. H. (2005). Comparing correlations. In *Contemporary Psychometrics: A festschrift for Roderick P. McDonald.* (Edited by A. Maydeu-Olivares and J. J. McArdle), Lawrence Erlbaum Associates: Mahwah, New Jersey.

Wager, T. D., Jonides, J. and Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage* **22**, 1679-1693.

Wager, T. D., Jonides, J. and Smith, E. E. (2006). Individual differences in multiple types of shifting attention. *Memory & Cognition.* **34**, 1730-1743.

Wager, T. D., Jonides, J., Smith, E. E. and Nichols, T. E. (2005a). Towards a taxonomy of attention-shifting: Individual differences in fMRI during multiple shift types. *Cogn. Affect. Behav. Neurosci.* **5**, 127-143.

Wager, T. D., Keller, M. C., Lacey, S. C., and Jonides, J. (2005b). Increased sensitivity in neuroimaging analyses using robust regression. *NeuroImage* **26**, 99-113.

Wager, T. D. and Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* **3**, 255-274.

Wager, T. D., Sylvester, C. Y., Lacey, S. C., Nee, D. E., Franklin, M. and Jonides, J. (2005c). Common and unique components of response inhibition revealed by fMRI. *NeuroImage* **27**, 323-340.

Ward, G., Roberts, M. J. and Phillips, L. H. (2001). Task-switching costs, Stroop-costs, and executive control: A correlational study. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **54A**, 491-511.

Woolrich, M. W., Ripley, B. D., Brady, J. M. and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modelling of fMRI data. *NeuroImage* **14**, 1370-1386.

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., and Smith, S.M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage* **21**, 1732-1747.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fMRI time series revisited – again. *NeuroImage* **2**, 173-181.

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., and Evans, A.C. (2002). A general statistical analysis for fMRI data. *NeuroImage* **15**, 1-15.

Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700 108, India.

E-mail: sumitra@isical.ac.in

Department of Psychology, Columbia University, 1190 Amsterdam Ave, New York, NY 10027, U.S.A.

E-mail: tor@paradox.psych.columbia.edu

GlaxoSmithKline, Clinical Imaging Centre, Imperial College, London, Hammersmith Hospital, Du Cane Road, London W12 0NN, U.K.

E-mail: thomas.e.nichols@gsk.com

(Received April 2007; accepted October 2007)